# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB05/000244

International filing date: 26 January 2005 (26.01.2005)

Document type: Certified copy of priority document

Document details: Country/Office: GB
Number: 0401639.0
Filing date: 26 January 2004 (26.01.2004)

Date of receipt at the International Bureau: 02 March 2005 (02.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

PCT/GB 2005/000244

PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS
The Patent Office

INVESTOR IN PEOPLE

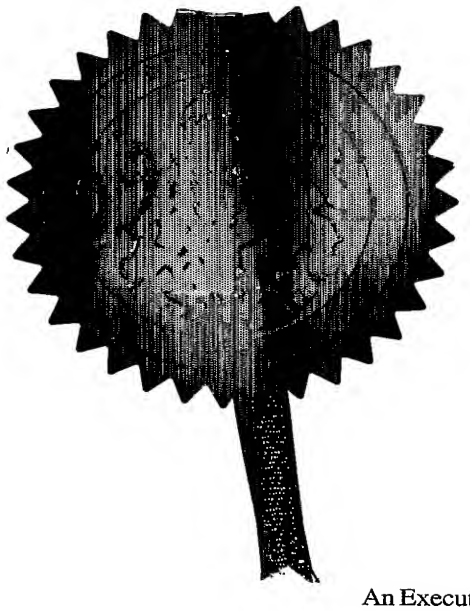
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed Stephen Hordley

Dated 11 February 2005

**Patents Form 1/77**

[redacted]ct 1977
(Rule 16)

The Patent Office

27JAN04 E868321-1 D00056
P01/7700 0.00-0401639.0 CHEQUE

THE PATENT OFFICE
L
26 JAN 2004
LONDON

**The Patent Office**

Cardiff Road
Newport
South Wales
NP9 1RH

# Request for grant of a patent

***(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)***

1. Your reference P36334GB KVC

2. Patent application number *(The Patent Office will fill in this part)*
~~26 JAN 2004~~ 0401639.0

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Isis Innovation Ltd
Ewert House, Ewert Place
Summertown, Oxfordshire
OX2 7SG

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation

United Kingdom

3998564003

4. Title of the invention

Molecular Analysis

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Kilburn & Strode
20 Red Lion Street
London
WC1R 4PJ

Patents ADP number *(if you know it)*

125001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

| Country | Priority application number *(if you know it)* | Date of filing *(day / month / year)* |
|---|---|---|

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

| Number of earlier application | Date of filing *(day / month / year)* |
|---|---|

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

*a) any applicant named in part 3 is not an inventor, or*
*b) there is an inventor who is not named as an applicant, or*
*c) any named applicant is a corporate body.*
*See note (d))*

YES

**Patents Form 1/77**

[redacted]er the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

| | |
|---|---|
| Continuation sheets of this form | |
| Description | 6 |
| Claim *(s)* | 2 |
| Abstract | 0 |
| Drawing *(s)* | 0 |

10. If you are also filing any of the following, state how many against each item.

| | |
|---|---|
| Priority documents | 0 |
| Translations of priority documents | 0 |
| Statement of inventorship and right to grant of a patent *(Patents Form 7/77)* | 0 |
| Request for preliminary examination and search *(Patents Form 9/77)* | 0 |
| Request for substantive examination *(Patents Form 10/77)* | 0 |
| Any other documents *(please specify)* | 0 |

11. I/We request the grant of a patent on the basis of this application.

Signature Kilburn & Strode Date 26/01/2004

12. Name and daytime telephone number of person to contact in the United Kingdom

Tel: 020 7539 4200

**Warning**

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.*

**Notes**

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

# Molecular Analysis

The present invention relates to a method of determining the sequence or occurrence frequency of a number of variable gene inserts from a gene library, wherein each variable gene insert is flanked 5' and 3' by known sequences, the method comprising; conducting polymerase chain reaction to amplify the variable gene inserts to produce components of a mixed PCR product, ligating the components of the mixed PCR product to produce a concatenated sequence and sequencing or determining the occurrence of the gene inserts in the concatenated sequence.

## Background to the invention

Peptide phage display is a prototypical version of directed *in vitro* molecular evolution of a large combinatorial library by sequential rounds of physical selection and enrichment. Peptide phage display selection methods have established themselves as powerful tools for the identification of short linear peptide mimetics of many ligand classes. Variant techniques have also been developed extending this methodology to selection from large libraries of oligonucleotides (either random or constrained), translation-arrested ribosomes, phage-displayed binding proteins (of which single chain fragments of immunoglobulin is the largest group) and so on.

A major limitation of all these methods is that the complexity and composition of the selection-evolved sublibrary is assessed by analysis of a very small sample drawn at random from this sublibrary. In consequence, enrichment is deemed to have been achieved when a very few, or more usually one, sequence dominates the sample. This may be acceptable when the evolution is directed by a simple target with one or very few binding sites, but severely limits the method if the target is complex.

The outcome of repeated rounds of selection from a large random peptide phage display library (typically beginning with $>10^9$ different phage) is a reduced complexity sub-library enriched for sequences showing specific affinity for the selection matrix. Typically such a sub-library may contain $10^3$-$10^4$ different phage,

but in all published studies the outcome of phage panning is assessed by sequencing <20 independent clones.

The problems inherent in directed evolution with a complex target have already been recognised in the field; a theoretical analysis has been presented (Vant Hull et al (1998) J Mol Biol 278 597-597) and a partial practical solution suggested (Messmer et al (2000) J Mol Biol 296 821-832). This "iterative panning" solution relies on completing multiple rounds of directed evolution until a single 'winning' sequence emerges. This sequence is then prepared as a synthetic peptide and used in the blocking solution during a repeat of the whole experiment (ie a further set of selections on new target material). Binding of the first class of peptides is now blocked and a second 'winning' sequence is selected. This process can be continued indefinitely, each entire round generating one new binding sequence. The method is very slow, very expensive and probably impossible for many key complex biological target materials, since it relies on a large supply of functionally homogenous target material (tumour tissue, infected cells etc) for many rounds of selection.

The present invention addresses the problems identified in the prior art.

The present invention provides a method of determining the sequence and/or occurrence frequency of a number of variable gene inserts from a gene library, wherein each variable gene insert is flanked 5' and 3' by known sequences, the method comprising;

conducting polymerase chain reaction to amplify the variable gene inserts to produce components of a mixed PCR product;

ligating the components of the mixed PCR product to produce a concatenated sequence; and

sequencing or determining the occurrence of the gene inserts in the concatenated sequence.

In accordance with the invention, the method can be used for determining the sequence of a number of variable gene inserts or for determining the occurrence frequency of a number of gene inserts from a gene library. Preferably the gene library is a peptide phage display library.

In the present invention, the polymerase chain reaction is conducted using primers complementary to the known sequences 5' and 3' to the variable inserts. Further, between the step of ligating the components of the mixed PCR product to produce a concatenated sequence and sequencing or determining the occurrence frequency of the gene inserts, it is preferable to subclone the size-selected concatenated products into a convenient vector for production of plasmid DNA suitable for automated sequencing.

The present invention relates to a library (of any size, including a large library) which may have been selected or evolved by cycles of physical selection and amplification to generate a sub-library whose gene inserts encode sequences that share some desired binding property.

The invention provides a simple and economical way of sequencing the relevant variable parts of the gene encoding the phage code protein from a large number (preferably all) of the phage in the selected sub-library. This is in contrast to the prior art which only determined the sequence of a tiny and potentially unrepresentative sub-set of the library. Furthermore, the present invention is a large-scale unbiased analysis without plaque selection or phage DNA purification from selected plaques. The method is achieved by using a polymerase chain reaction with unique primers lying just 5' and just 3' to the variable insert in the gene encoding phage coat protein. The reaction is carried out on pooled phage DNA isolated from an aliquot of the library without plaque purification and therefore contains proportional representation amplification of all variable regions in the library or sub-library.

The benefit of the present invention is that each variable region will have an abundance in the double strand DNA product that is proportional to the abundance of that insert sequence amongst the phage in the selected sub-library. When the components of the mixed PCR product have been prepared, they are ligated to produce a concatenated sequence. It may be useful or preferable to digest the components of the mixed PCR product with a very infrequently cutting restriction endonuclease before ligation to produce concatenated sequences. Following production of the concatenated sequences, they may preferably be size selected to around 1.5kb before being cloned into a convenient plasmid. Subsequent sequencing of such inserts generates greater than 30 variable insert sequences per sequencing lane.

Known software that is used to automatically strip the joining sequences out of continuous DNA sequence to identify and then tabulate the di-tags during serial analysis of ligand-selected peptide display sub-libraries can be used to generate abundance histograms for all the insert sequences identified.

The present invention allows the easy analysis of many (often all) of the variable inserts present in the gene library population. This permits the evolution of the selection process to be followed much more accurately. It also ensures that consensus matrix-binding sequences are identified both earlier and more accurately. Also important is that the method of the present invention overcomes the problems of clonal dominance due to the emergence of a single family of binding sequences which prevents analysis of interactions on complex matrices.

The present invention allows the rapid and complete identification of all linear or cysteine cyclised peptides that exhibit a specific behaviour permitting gene selection (either positive or negative). It is also applicable to the classification of all antibody epitopes in a complex humoral response to a pathogen.

The source of the variable gene inserts in the library is nucleic acid. It may be one or more selected from the source of a bacterium, virus, peptide-mimetic, immunoglobulin or cells, including infected cells and tumour cells.

The specific behaviour permitting gene selection (also described herein as a specific characteristic permitting gene selection) may be any, including the fact that the gene encodes a particular protein which binds to another protein in question. Alternatively, the gene may encode a protein sequence which only occurs in one state of tissue in comparison with the same tissue in a different state. For example, normal versus tumour tissue, infected versus non-infected tissue, wild type versus mutant, healthy versus oxidatively damaged, healthy versus ischaemic, or occurrence during a particular time zone which is absent at an alternative time zone.

The essence of the proposed invention is a method based on concatenation of short PCR products for efficient sequencing; this permits the analysis of hundreds if not thousands of sequences corresponding to peptides selected at each round of a target-directed evolution from a large combinatorial library. In its simplest form this method reveals multiple sequence families as they are enriched by selection; a single series of enrichment experiments generates frequency histograms for all emergent classes of selected sequence. In this way, multiple binding sites on a complex target are identified without using the time-consuming and expensive iterative panning approach.

The present invention can be utilised in various ways. For example, differential panning on two states (normal versus tumour tissue; infected versus non-infected; wild-type versus mutant; healthy versus oxidatively damaged; healthy versus ischaemic, etc) together with frequency histogram generation on large insert numbers at each round of panning offers a new type of information. The frequency histograms of the two independent panning experiments are compared (in a manner analogous to comparing two SAGE tag profiles, or the microarray binding data from the mRNA samples). This identifies peptide binders that are state independent (ie lying along the diagonal on the two state plot) as well as binders that are enriched in one state or the

other. This already enriches the data obtained very substantially. However, by adding a third dimension that identifies the time of appearance of a given sequence during the multiple rounds of panning an entirely new type of information emerges. It is now possible to perform cluster analysis on points that lie off the diagonal within this 3-D space to identify groups of weak signals that together offer a discriminant measure between the two states. To take a specific example, this approach could be used to search for small groups of peptide mimetics that can together discriminate between normal and tumour tissue in a way that could not be achieved by analysis of binding of any single peptide.

## Claims

1. A method of determining the sequence or occurrence frequency of a number of variable gene inserts from a gene library, wherein each variable gene insert is flanked 5' and 3' adjacent known sequences, the method comprising;

conducting polymerase chain reaction to amplify the variable gene inserts to produce components of a mixed PCR product;

ligating the components of the mixed PCR product to produce a concatenated sequence; and

sequencing or determining the occurrence of the gene inserts in the concatenated sequence.

2. A method as claimed in claim 1 wherein the gene library is a peptide phage display library.

3. A method as claimed in claim 1 or claim 2, wherein the components of the mixed PCR product are digested with a restriction endonuclease before ligation to produce the concatenated sequence.

4. A method as claimed in any one of claims 1 to 3, wherein the concatenated sequence is cloned into a plasmid before sequencing.

5. A method as claimed in claim 4, wherein the concatenated sequence is size selected to around 1.5 kilobases in length before cloning into the plasmid.

6. A method as claimed in any one of claims 2 to 5, wherein the number of variable gene inserts are from phage which exhibit a specific characteristic.

7. A method as claimed in claim 6, wherein the specific characteristic is one or

more of protein binding or occurrence only in one state of tissue in a comparison with the same tissue in a different state.

8. A method as claimed in any one of claims 2 to 5, wherein the number of variable gene inserts are from phage which do not exhibit a specific characteristic.

9. A method as claimed in any one of claims 1 to 9, wherein the source of the variable gene inserts is nucleic and one or more of a bacterium, virus, peptide-mimetic immunoglobulin or cells including infected cells and tumour cells.

10. A method as claimed in any one of claims 1 to 9, wherein frequency analysis of sequential rounds of selection of variable gene inserts from a gene library is used to perform discriminant analysis of the states of the selection method.

11. A determination of the sequence or occurrence frequency of a number of variable gene inserts from a gene library, obtained by a method according to any one of claims 1 to 10.